Weighting-Like Diagnostics for Nonlinear Bayesian Hierarchical Models

Ryan Giordano, Alice Cima, Erin Hartman, Jared Murray, Avi Feller

October 2025 Stanford Berkeley Joint Colloquium

Are US non-voters becoming more Republican?

Blue Rose research says yes:

"Politically disengaged voters have become much more Republican, and because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate."

(Blue Rose Research 2024)
(major professional pollsters)

On Data and Democracy **says no:**

"Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available."

(Bonica et al. 2025)
(major professional researchers)

Are US non-voters becoming more Republican?

Blue Rose research says yes:

“Politically disengaged voters have become much more Republican, and because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate.”

(Blue Rose Research 2024)
(major professional pollsters)

On Data and Democracy **says no:**

“Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available.”

(Bonica et al. 2025)
(major professional researchers)

- The problem is very hard (it’s difficult to accurately poll non–voters)
- Different data sources
- ⋆⋆⋆ **Different statistical methods**
  - Blue Rose uses Bayesian hierarchical modeling (MrP)
  - On Data and Democracy is using calibration weighting (CW)

Are US non-voters becoming more Republican?

Blue Rose research says yes:

"Politically disengaged voters have become much more Republican, and because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate."

(Blue Rose Research 2024)
(major professional pollsters)

***On Data and Democracy* says no:**

"Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available."

(Bonica et al. 2025)
(major professional researchers)

- The problem is very hard (it's difficult to accurately poll non–voters)
- Different data sources
- ⋆⋆⋆ **Different statistical methods**
  - Blue Rose uses Bayesian hierarchical modeling (MrP)
  - On Data and Democracy is using calibration weighting (CW)

Our contribution

We define "MrP local equivalent weights" (MrPlew) that:

- Are easily computable from MCMC draws and standard software, and
- Provide MrP versions of key weighting estimator diagnostics.

⇒ **MrPlew provides direct comparisons between MrP and calibration weighting.**

Outline

- Introduce the statisical problem
  - Contrast calibration weighting and MrP
  - Prior work: Equivalent weights for linear models
  - Equivalent weights and implicit weights for non–linear models
  - Our task: Rigorously justify using locally equivalent weights for diagnostics

Outline

- Introduce the statisical problem
  - Contrast calibration weighting and MrP
  - Prior work: Equivalent weights for linear models
  - Equivalent weights and implicit weights for non–linear models
  - Our task: Rigorously justify using locally equivalent weights for diagnostics
- Locally equivalent weights for frequentist variance estimation

Outline

- Introduce the statisical problem
  - Contrast calibration weighting and MrP
  - Prior work: Equivalent weights for linear models
  - Equivalent weights and implicit weights for non–linear models
  - Our task: Rigorously justify using locally equivalent weights for diagnostics
- Locally equivalent weights for frequentist variance estimation
- Locally equivalent weights for covariate balance
  - Describe classical covariate balance
  - Introduce a MrPlew “local empirical consistency check”
  - Theoretical support
  - Examples of real-world results

Outline

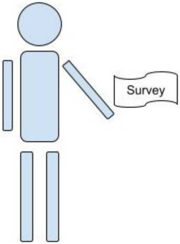
- Introduce the statisical problem
  - Contrast calibration weighting and MrP
  - Prior work: Equivalent weights for linear models
  - Equivalent weights and implicit weights for non–linear models
  - Our task: Rigorously justify using locally equivalent weights for diagnostics
- Locally equivalent weights for frequentist variance estimation
- Locally equivalent weights for covariate balance
  - Describe classical covariate balance
  - Introduce a MrPlew "local empirical consistency check"
  - Theoretical support
  - Examples of real-world results
- Other directions
  - High–level restatement of the logic of our procedure
  - Local versions of other common diagnostics for linear estimators
  - Ongoing and future work

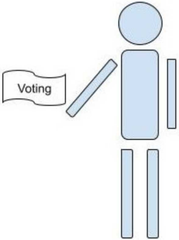
The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to "do you support Trump")

We want the average response in a target population, in which we observe only covariates.





Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, N_S$ | Observe $\mathbf{x}_j$ for $j = 1, \ldots, N_T$

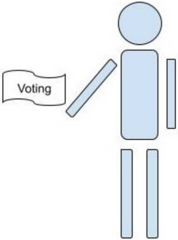
[1] Photo copyright: Mark Taylor / `naturepl.com`

The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to "do you support Trump")

We want the average response in a target population, in which we observe only covariates.



Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, N_S$ | Observe $\mathbf{x}_j$ for $j = 1, \ldots, N_T$

The problem is that the populations may be very different, maybe leading to bias. [1]

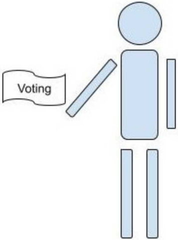
[1] Photo copyright: Mark Taylor / `naturepl.com`

The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to “do you support Trump”)

We want the average response in a target population, in which we observe only covariates.



Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, N_S$ | Observe $\mathbf{x}_j$ for $j = 1, \ldots, N_T$

The problem is that the populations may be very different, maybe leading to bias. [1]

How can we use the covariates to say something about the target responses?

[1] Photo copyright: Mark Taylor / `naturepl.com`

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j. Let $Y_{\mathcal{S}} = \{y_1, \ldots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j. Let $Y_{\mathcal{S}} = \{y_1, \ldots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

▶ Choose "calibration weights" w_i
using only the regressors $\mathbf{x}$
(e.g. raking weights)

Bayesian hierarchical modeling (MrP)

▶ Choose $\mathbb{E}\left[y|\mathbf{x}, \theta\right] = m(\theta^\intercal \mathbf{x})$,
choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$
(e.g. Hierarchical logistic regression)

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j. Let $Y_{\mathcal{S}} = \{y_1, \ldots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- Choose "calibration weights" w_i using only the regressors $\mathbf{x}$ (e.g. raking weights)
- Take $\hat{\mu}^{\text{WGT}}(Y_{\mathcal{S}}) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$

Bayesian hierarchical modeling (MrP)

- Choose $\mathbb{E}\left[y|\mathbf{x}, \theta\right] = m(\theta^\intercal \mathbf{x})$, choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$ (e.g. Hierarchical logistic regression)
- Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[y|\mathbf{x}_j\right]$ and $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j. Let $Y_\mathcal{S} = \{y_1, \ldots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- Choose "calibration weights" w_i using only the regressors $\mathbf{x}$ (e.g. raking weights)
- Take $\hat{\mu}^{\mathrm{WGT}}(Y_\mathcal{S}) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$
- Dependence on y_i is clear

Bayesian hierarchical modeling (MrP)

- Choose $\mathbb{E}\left[y|\mathbf{x}, \theta\right] = m(\theta^\intercal \mathbf{x})$, choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$ (e.g. Hierarchical logistic regression)
- Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[y|\mathbf{x}_j\right]$ and $\hat{\mu}^{\mathrm{MrP}}(Y_\mathcal{S}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$
- Dependence on y_i very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j. Let $Y_{\mathcal{S}} = \{y_1, \ldots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

▶ Choose "calibration weights" w_i using only the regressors $\mathbf{x}$ (e.g. raking weights)

▶ Take $\hat{\mu}^{\mathrm{WGT}}(Y_{\mathcal{S}}) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$

▶ Dependence on y_i is clear

▶ Weights give interpretable diagnostics:

- Frequentist variability
- Regressor balance
- Partial pooling

Bayesian hierarchical modeling (MrP)

▶ Choose $\mathbb{E}\left[y|\mathbf{x}, \theta\right] = m(\theta^\intercal \mathbf{x})$, choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$ (e.g. Hierarchical logistic regression)

▶ Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[y|\mathbf{x}_j\right]$ and $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$

▶ Dependence on y_i very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)

▶ **Black box**

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j. Let $Y_\mathcal{S} = \{y_1, \ldots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

▶ Choose "calibration weights" w_i using only the regressors $\mathbf{x}$ (e.g. raking weights)

▶ Take $\hat{\mu}^{\mathrm{WGT}}(Y_\mathcal{S}) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$

▶ Dependence on y_i is clear

▶ Weights give interpretable diagnostics:

- Frequentist variability
- Regressor balance
- Partial pooling

Bayesian hierarchical modeling (MrP)

▶ Choose $\mathbb{E}\left[y|\mathbf{x}, \theta\right] = m(\theta^\intercal \mathbf{x})$, choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$ (e.g. Hierarchical logistic regression)

▶ Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[y|\mathbf{x}_j\right]$ and $\hat{\mu}^{\mathrm{MrP}}(Y_\mathcal{S}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$

▶ Dependence on y_i very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)

▶ **Black box**

$\leftarrow$ Today, we'll open the box and provide MrP analogues of all these diagnostics

Prior work: Equivalent weights for linear models

Gelman (2007b) observes that MrP is a weighting estimator when $\hat{y}$ is computed with OLS:

$$\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T}\sum_{j=1}^{N_T}\hat{y}_j = \frac{1}{N_T}\sum_{j=1}^{N_T}\underbrace{\mathbf{x}_j^{\intercal}\hat{\theta}}_{\text{Linear in } Y_{\mathcal{S}}}$$

Most existing literature on comparing weighting and MrP focus on such linear models. [2]

[2] For example, Gelman (2007b), B., F., and H. (2021), and Chattopadhyay and Zubizarreta (2023).

Prior work: Equivalent weights for linear models

Gelman (2007b) observes that MrP is a weighting estimator when $\hat{y}$ is computed with OLS:

$$\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T}\sum_{j=1}^{N_T} \hat{y}_j = \frac{1}{N_T}\sum_{j=1}^{N_T} \underbrace{\mathbf{x}_j^{\intercal}\hat{\theta}}_{\text{Linear in } Y_{\mathcal{S}}}$$

Most existing literature on comparing weighting and MrP focus on such linear models. [2]

But what if you use a non–linear link function? Or a hierarchical model?

> *"It would also be desirable to use nonlinear methods ... but then it would seem difficult to construct even approximately equivalent weights. Weighting and fully nonlinear models would seem to be completely incompatible methods."* — *(Gelman 2007a)*

[2] For example, Gelman (2007b), B., F., and H. (2021), and Chattopadhyay and Zubizarreta (2023).

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

The map from $Y_{\mathcal{S}} \mapsto m(\mathbf{x}_j^\intercal \hat{\theta})$ is *inherently nonlinear*.

But *some sample averages* of $m(\mathbf{x}_j^\intercal \hat{\theta})$ can be approximately linear.

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal\theta) = \mathrm{Logistic}(\mathbf{x}^\intercal\theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T}\sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal\hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal\mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$\begin{aligned} \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) \\ &\approx \int m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \end{aligned}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$
\begin{aligned}
\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) \\
&\approx \int m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\
&= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x})\text{)}
\end{aligned}
$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$
\begin{aligned}
\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) \\
&\approx \int m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\
&= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x})\text{)} \\
&\approx \int (\alpha^\intercal \mathbf{x})\, m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(By assumption)}
\end{aligned}
$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$
\begin{aligned}
\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) \\
&\approx \int m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\
&= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x})\text{)} \\
&\approx \int (\alpha^\intercal \mathbf{x})\, m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(By assumption)} \\
&\approx \alpha^\intercal \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{x}_i m(\mathbf{x}_i^\intercal \hat{\theta}) && \text{(Law of large numbers)}
\end{aligned}
$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$
\begin{aligned}
\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) \\
&\approx \int m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\
&= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x})\text{)} \\
&\approx \int (\alpha^\intercal \mathbf{x})\, m(\mathbf{x}^\intercal \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(By assumption)} \\
&\approx \alpha^\intercal \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{x}_i m(\mathbf{x}_i^\intercal \hat{\theta}) && \text{(Law of large numbers)} \\
&= \alpha^\intercal \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{x}_i y_i && \text{(Property of exponential family MLEs)}
\end{aligned}
$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\mathrm{MrP}}}_{\alpha^\intercal \mathbf{x}_i} y_i + \text{Small error}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\text{MrP}}}_{\alpha^\intercal \mathbf{x}_i} y_i + \text{Small error}$$

But what are the weights? We don't observe $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})}$, so can't estimate α directly.

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\mathrm{MrP}}}_{\alpha^\intercal \mathbf{x}_i} y_i + \text{Small error}$$

Key idea (informal)

If $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ is approximately linear, then[3] $w_i^{\mathrm{MrP}} \approx N_S \frac{\partial \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$.

[3]For MLEs, $\frac{\partial \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$ is given by the implicit function theorem. (Krantz and Parks 2012; **G.**, Stephenson, et al. 2019)

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\intercal \mathbf{x}$ for some α. **Then MrP is a *approximately* a CW estimator.**

$$\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\intercal \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\text{MrP}}}_{\alpha^\intercal \mathbf{x}_i} y_i + \text{Small error}$$

Key idea (informal)

If $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ is approximately linear, then[3] $w_i^{\text{MrP}} \approx N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$.

Note: The derivatives w_i^{MrP} now have two potentially distinct interpretations:

- **Equivalent weights:** A characterization of $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ for diagnostics
- **Implicit weights:** An estimate of $\mathcal{P}_T(\mathbf{x})/\mathcal{P}_S(\mathbf{x})$

[3] For MLEs, $\frac{\partial \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$ is given by the implicit function theorem. (Krantz and Parks 2012; **G.**, Stephenson, et al. 2019)

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \mathrm{Logistic}(\mathbf{x}^\intercal \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\intercal \theta) \right]$.

No reason to think $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ is even approximately **globally** linear.

---

[4] Diaconis and Freedman 1986; Gustafson 1996; Efron 2015; **G.**, Broderick, and Jordan 2018.

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\intercal \theta) \right]$.

No reason to think $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ is even approximately **globally** linear.

But can still compute and analyze $w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$ using Bayesian sensitivity analysis![4]

MrP weights for MCMC

$$w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})}{\partial y_i} = N_S \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\text{Cov}_{\mathcal{P}(\theta|\text{Survey data})} \left(m(\mathbf{x}_j^\intercal \theta), \theta^\intercal \mathbf{x}_i \right)}_{\text{Can estimate without rerunning MCMC!}}$$

[4] Diaconis and Freedman 1986; Gustafson 1996; Efron 2015; **G.**, Broderick, and Jordan 2018.

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\intercal \theta) \right]$.

No reason to think $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ is even approximately **globally** linear.

But can still compute and analyze $w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$ using Bayesian sensitivity analysis![4]

MrP weights for MCMC

$$w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})}{\partial y_i} = N_S \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\text{Cov}_{\mathcal{P}(\theta|\text{Survey data})} \left(m(\mathbf{x}_j^\intercal \theta), \theta^\intercal \mathbf{x}_i \right)}_{\text{Can estimate without rerunning MCMC!}}$$

The derivatives w_i^{MrP} *again* have two potentially distinct interpretations:

- **Locally equivalent weights:** A characterization of $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ for diagnostics
- **Locally implicit weights:** An estimate of $\mathcal{P}_T(\mathbf{x})/\mathcal{P}_S(\mathbf{x})$

[4] Diaconis and Freedman 1986; Gustafson 1996; Efron 2015; **G.**, Broderick, and Jordan 2018.

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal\theta) = \mathrm{Logistic}(\mathbf{x}^\intercal\theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T}\sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[m(\mathbf{x}_j^\intercal\theta)\right]$.

No reason to think $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ is even approximately **globally** linear.

But can still compute and analyze $w_i^{\mathrm{MrP}} := N_S \frac{\partial \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})}{\partial y_i}$ using Bayesian sensitivity analysis![4]

MrP weights for MCMC

$$w_i^{\mathrm{MrP}} := N_S \frac{\partial \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})}{\partial y_i} = N_S \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\mathrm{Cov}_{\mathcal{P}(\theta|\text{Survey data})}\left(m(\mathbf{x}_j^\intercal\theta), \theta^\intercal \mathbf{x}_i\right)}_{\text{Can estimate without rerunning MCMC!}}$$

The derivatives w_i^{MrP} *again* have two potentially distinct interpretations:

- **Locally equivalent weights:** A characterization of $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ for diagnostics
- **Locally implicit weights:** An estimate of $\mathcal{P}_T(\mathbf{x})/\mathcal{P}_S(\mathbf{x})$

This talk will focus only on locally equivalent weights. (Implicit weights is ongoing work!)

[4] Diaconis and Freedman 1986; Gustafson 1996; Efron 2015; **G.**, Broderick, and Jordan 2018.

Locally equivalent weights for hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\intercal \theta) = \text{Logistic}(\mathbf{x}^\intercal \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\intercal \theta) \right]$.

MrP locally equivalent weights (MrPlew)

For new data $\tilde{Y}_{\mathcal{S}}$, form a **MrP locally equivalent weighting**:

$$\hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) \approx \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) + \sum_{i=1}^{N_{\mathcal{S}}} w_i^{\text{MrP}} (\tilde{y}_i - y_i)$$

Our task is to rigorously show that even such local weights can be meaningfully used diagnostically in the same ways we use global weights.

Real Data: Marital Name Change Survey

Analysis of changing names after marriage[5].

- **Target population:** ACS survey of US population 2017–2022[6]
- **Survey population:** Marital Name Change Survey (from Twitter)[7]
- **Respose:** Did the female partner keep their name after marriage?
- For regressors, use bins of age, education, state, and decade married.

Survey observations: $N_S = 4,364$

Target observations (rows): $N_T = 4,085,282$

Uncorrected survey mean: $\frac{1}{N_S} \sum_{i=1}^{N_S} y_i = 0.462$

Raking: $\hat{\mu}^{\mathrm{WGT}}(Y_{\mathcal{S}}) = 0.263$

MrP: $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = 0.288 \quad (\text{Post. sd} = 0.0169)$

[5] Based on Alexander (2019).

[6] Ruggles et al. 2024.

[7] Cohen (2019)

Real Data: Marital Name Change Survey

Analysis of changing names after marriage[5].

- **Target population:** ACS survey of US population 2017–2022[6]
- **Survey population:** Marital Name Change Survey (from Twitter)[7]
- **Respose:** Did the female partner keep their name after marriage?
- For regressors, use bins of age, education, state, and decade married.

Survey observations: $N_S = 4,364$

Target observations (rows): $N_T = 4,085,282$

Uncorrected survey mean: $\frac{1}{N_S}\sum_{i=1}^{N_S} y_i = 0.462$

Raking: $\hat{\mu}^{\mathrm{WGT}}(Y_{\mathcal{S}}) = 0.263$

MrP: $\hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) = 0.288$ $(\text{Post. sd} = 0.0169)$

[5] Based on Alexander (2019).
[6] Ruggles et al. 2024.
[7] Cohen (2019)

The weights can look very different!

Does this mean anything?

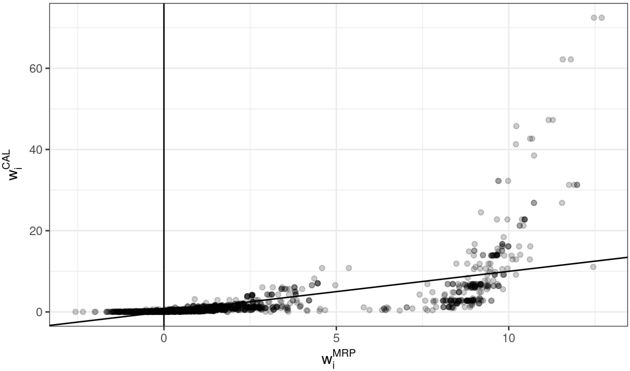


Figure 1: Comparison between raking and MrPlew weights for the Name Change dataset

The weights can look very different!

Does this mean anything?

Does the spread relate to frequentist variance?

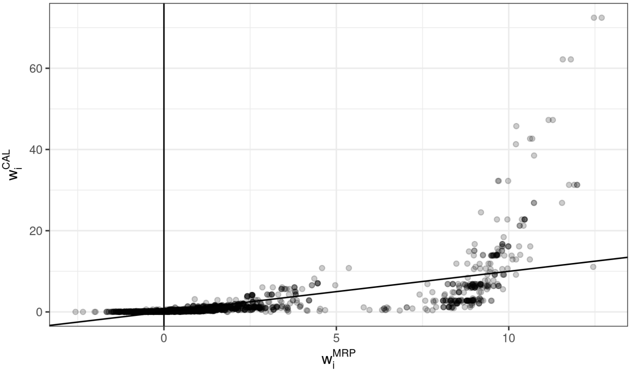


Figure 1: Comparison between raking and MrPlew weights for the Name Change dataset

Frequentist variance estimation

Let $\hat{\text{Var}}(\cdot)$ denote the sample variance.

Calibration weighting standard errors sketch: [8]

If we have $\hat{\mu}^{\text{WGT}}(Y_{\mathcal{S}}) = \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i$ and a consistent residual estimate ε_i, then

$$\hat{\text{Var}}(w_i \varepsilon_i) \approx \text{Var}\left(\sqrt{N_S}\hat{\mu}^{\text{WGT}}(Y_{\mathcal{S}})\right).$$

[8] E.g. , Deville, Särndal, and Sautory (1993) and Fuller (2011).

Frequentist variance estimation

Let $\hat{\text{Var}}(\cdot)$ denote the sample variance.

Calibration weighting standard errors sketch: [8]

If we have $\hat{\mu}^{\text{WGT}}(Y_{\mathcal{S}}) = \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i$ and a consistent residual estimate ε_i, then

$$\hat{\text{Var}}(w_i \varepsilon_i) \approx \text{Var}\left(\sqrt{N_S}\,\hat{\mu}^{\text{WGT}}(Y_{\mathcal{S}})\right).$$

MrPlew Standard error consistency theorem sketch (Our contribution): [9]

For Bayesian hierarchical logictic regression, define $\varepsilon_i = y_i - \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}\left[m(\mathbf{x}_i^\intercal \theta)\right]$.

We state mild conditions under which, as $N_S \to \infty$, for some μ_∞ and variance V,

$$\sqrt{N_S}\left(\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) - \mu_\infty\right) \to \mathcal{N}(0, V) \quad \text{and}$$
$$\hat{\text{Var}}\left(w_i^{\text{MrP}} \varepsilon_i\right) \to V.$$

The use of w_i^{MrP} is analogous to the use of w_i for frequentist variance estimation.

[8] E.g. , Deville, Särndal, and Sautory (1993) and Fuller (2011).

[9] This is essentially a corollary of our earlier work on the Bayesian infinitesimal jackknife. (**G.** and Broderick 2024)

Standard error estimation experiment

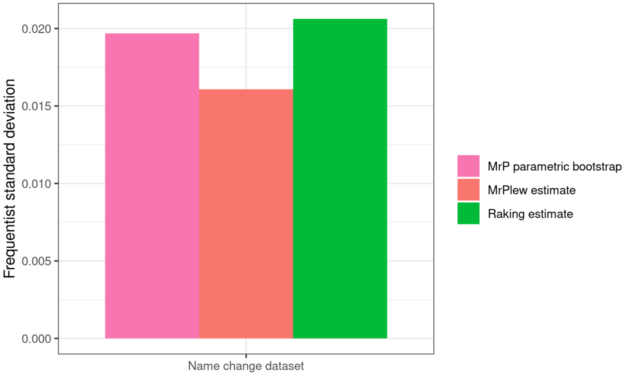


Figure 2: Frequentist standard deviation estimates

Standard error estimation experiment

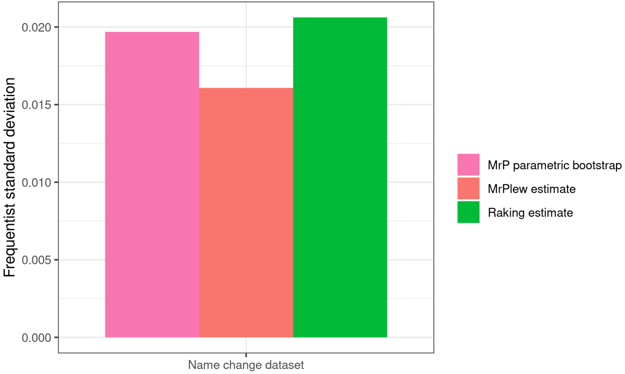


Figure 2: Frequentist standard deviation estimates

Running fifty MCMC parametric bootstraps: ≈ 79 hours

Computing approximate weights: 16 seconds

Other uses

Does this mean anything?

Yes: The “spread” relates to frequentist variance just as in weighting estimators.

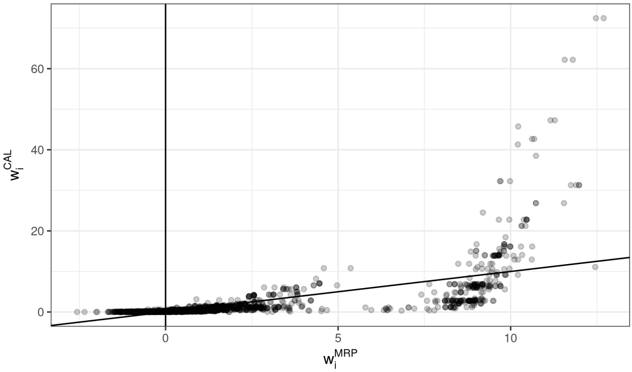


Figure 3: Comparison between raking and MrPlew weights for the Name Change dataset

Does this mean anything?

Yes: The "spread" relates to frequentist variance just as in weighting estimators.

What about covariate balance?

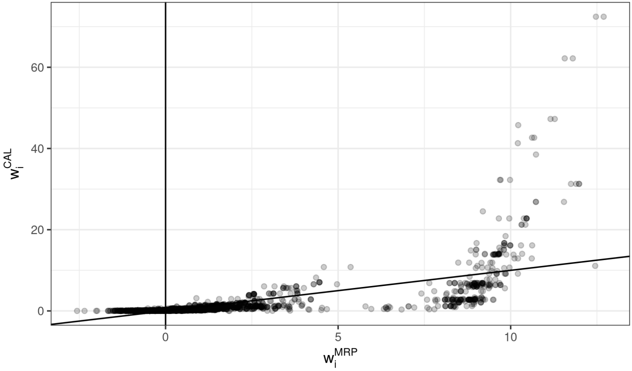


Figure 3: Comparison between raking and MrPlew weights for the Name Change dataset

Introduction to covariate balance: What are we weighting for?[10]

$$\text{Target average response} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j.

[10] Pun borrowed from Solon, Haider, and Wooldridge (2015)

Introduction to covariate balance: What are we weighting for?[10]

$$\text{Target average response} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j. But we can check whether:

$$\frac{1}{N_T}\sum_{j=1}^{N_T} \mathbf{x}_j \overset{\text{check}}{=} \frac{1}{N_S}\sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Weights that pass this check satisfy "covariate balance" for **x**.

[10] Pun borrowed from Solon, Haider, and Wooldridge (2015)

Introduction to covariate balance: What are we weighting for?[10]

$$\text{Target average response} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j. But we can check whether:

$$\frac{1}{N_T}\sum_{j=1}^{N_T} \mathbf{x}_j \overset{\text{check}}{=} \frac{1}{N_S}\sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Weights that pass this check satisfy "covariate balance" for $\mathbf{x}$.

You can check covariate balance for any weighting estimator, and any function $f(\mathbf{x})$.

Recall that **raking calibration weights** aim to exactly balance some set of regressors.

[10] Pun borrowed from Solon, Haider, and Wooldridge (2015)

Balance checks as local consistency

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Key idea: Define a data perturbation that captures this intuition.

Balance checks as local consistency

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (informal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the change this is supposed to induce in the target population.

Covariate balance checks whether our estimators produce the same change.

Balance checks as local consistency

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (formal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{N_T}\sum_{j=1}^{N_T}\left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{Y}_{\mathcal{S}}, Y_{\mathcal{S}}$:

$$\underbrace{\hat{\mu}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}(Y_{\mathcal{S}})}_{\text{Replace weighted averages with changes in an estimator}} \overset{\textbf{check}}{\approx} \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j).$$

Balance checks as local consistency

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (formal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{N_T}\sum_{j=1}^{N_T}\left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{Y}_{\mathcal{S}}, Y_{\mathcal{S}}$:

$$\underbrace{\hat{\mu}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}(Y_{\mathcal{S}})}_{\substack{\text{Replace weighted averages} \\ \text{with changes in an estimator}}} \overset{\textbf{check}}{\approx} \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j).$$

When $\hat{\mu}(\cdot) = \hat{\mu}^{\mathrm{WGT}}(\cdot)$, BISC recovers the standard covariate balance check.

We will study $\hat{\mu}(\cdot) = \hat{\mu}^{\mathrm{MrP}}(\cdot)$.

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$.

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$.

Problem: $\hat{\mu}^{\mathrm{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- MCMC output is noisy, and $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ may be small.

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$.

Problem: $\hat{\mu}^{\mathrm{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- MCMC output is noisy, and $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ may be small.

Solution: Use our local approximation, MrPlew!

Balance informed sensitivity check with MrPlew:

For a wide set of judiciously chosen $f(\cdot)$, check

$$\begin{aligned}
\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) &\approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i^{\mathrm{MrP}}(\tilde{y}_i - y_i) \\
&\approx \underbrace{\delta \frac{1}{N_S}\sum_{i=1}^{N_S} w_i^{\mathrm{MrP}} f(\mathbf{x}_i) \overset{\textbf{check}}{\approx} \delta \frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j)}_{\text{What you actually check}}.
\end{aligned}$$

When is the local approximation accurate?

BISC Theorem: (sketch)

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = \text{Small}$$

[11] $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[12] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

When is the local approximation accurate?

BISC Theorem: (sketch)

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2)$$

[11] $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[12] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

When is the local approximation accurate?

BISC Theorem: (sketch)

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \to \infty$$

[11] $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[12] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

When is the local approximation accurate?

BISC Theorem: (sketch)

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\sup_{f \in \mathcal{F}} \left| \hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \to \infty$$

...for a very broad class of $\mathcal{F}$. [11]

Uniformity justifies searching for "imbalanced" f.

[11] $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[12] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

When is the local approximation accurate?

BISC Theorem: (sketch)

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\sup_{f \in \mathcal{F}} \left| \hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \to \infty$$

...for a very broad class of $\mathcal{F}$. [11]

Uniformity justifies searching for "imbalanced" f.

The uniformity result builds on our earlier work on uniform and finite–sample error bounds for Bernstein–von Mises theorem–like results[12].

[11] $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[12] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

Covariate balance for primary effects

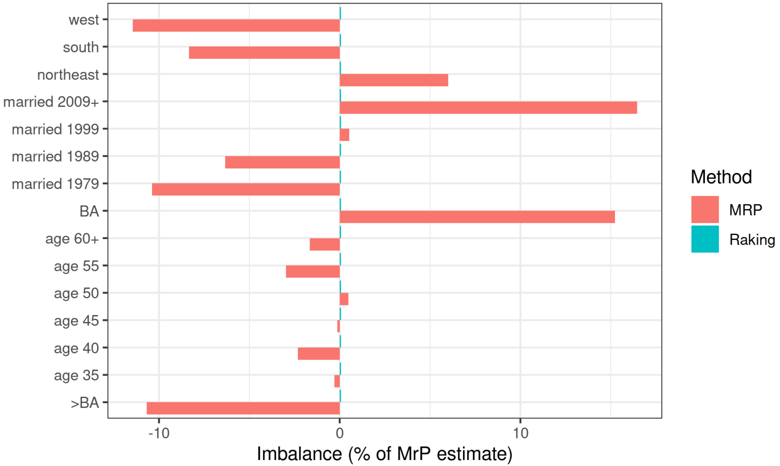


Figure 4: Imbalance plot for primary effects in the Name Change dataset

Covariate balance for interaction effects

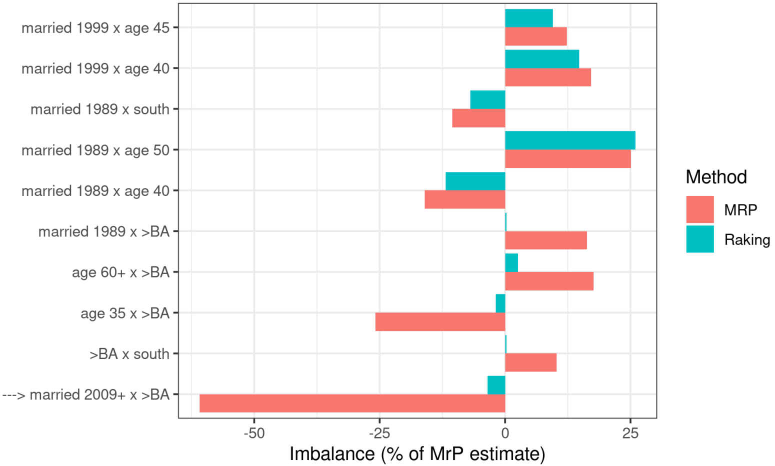


Figure 5: Imbalance plot for select interaction effects in the Name Change dataset

Predictions

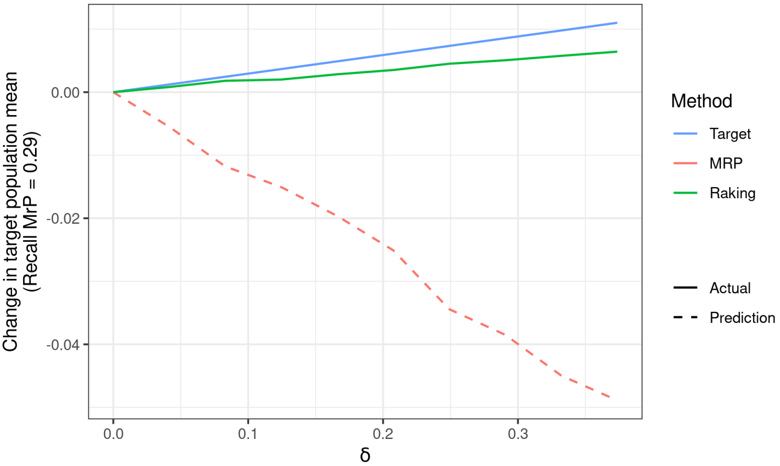


Figure 6: Predictions on binary data for the Name Change dataset

Predictions and actual MCMC results

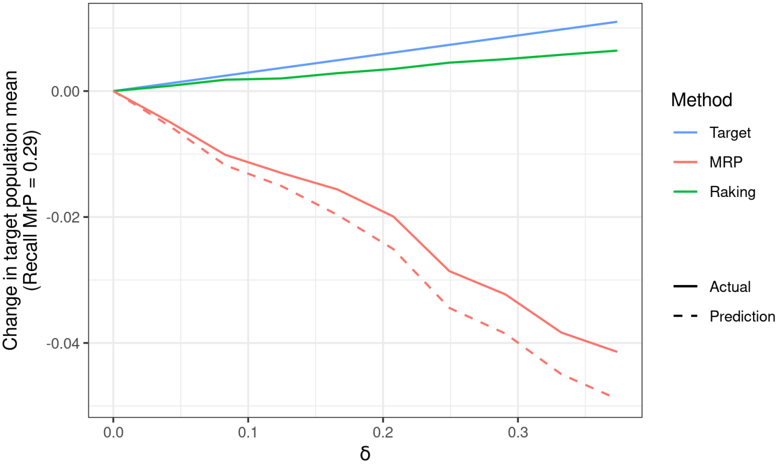


Figure 7: Predictions and refit on binary data for the Name Change dataset

Running ten MCMC refits: 10 hours Computing approximate weights: 16 seconds

Future work

Notice that there was no discussion of misspecification!

Calibration weights (typically) do not depend on $Y_{\mathcal{S}}$.

Future work

Notice that there was no discussion of misspecification!

Calibration weights (typically) do not depend on $Y_{\mathcal{S}}$.

But the high level idea can be extended much more widely:

1. Assume your initial model was accurate
2. Select some perturbation your model should be able to capture
3. Use local sensitivity to detect whether the change is what you expect
4. If the change is not what you expect, either (1) or (2) was wrong

Future work

Notice that there was no discussion of misspecification!

Calibration weights (typically) do not depend on $Y_{\mathcal{S}}$.

But the high level idea can be extended much more widely:

1. Assume your initial model was accurate
2. Select some perturbation your model should be able to capture
3. Use local sensitivity to detect whether the change is what you expect
4. If the change is not what you expect, either (1) or (2) was wrong

Checks of this form give generalized versions of many standard linear model diagnostics.

Some generalized diagnostics

Regression

Some generalized diagnostics

	Regression	General models

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogonous residuals	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y)$	

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogonous residuals	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y \mid \mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \overset{\text{check}}{=} \hat{\theta}(w)$

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogonous residuals	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y\vert\mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \overset{\text{check}}{=} \hat{\theta}(w)$
Fisher information	$\mathcal{I} :=$ Fisher information $\Sigma :=$ Score covariance $\mathcal{I}^{-1} \overset{\text{check}}{=} \Sigma$	

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogonous residuals	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y\vert\mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \overset{\text{check}}{=} \hat{\theta}(w)$
Fisher information	$\mathcal{I} :=$ Fisher information $\Sigma :=$ Score covariance $\mathcal{I}^{-1} \overset{\text{check}}{=} \Sigma$	$y \sim \mathcal{P}(y\vert\theta)$ $\tilde{y} \sim$ Importance sample y using $\tilde{w} = \dfrac{\mathcal{P}(y\vert\hat{\theta} + \delta)}{\mathcal{P}(y\vert\hat{\theta})}$ $\hat{\theta}(\tilde{w}) \overset{\text{check}}{=} \hat{\theta}(1) + \delta$

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\intercal \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogonous residuals	$y = \theta^\intercal \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \overset{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y\vert\mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \overset{\text{check}}{=} \hat{\theta}(w)$
Fisher information	$\mathcal{I} :=$ Fisher information $\Sigma :=$ Score covariance $\mathcal{I}^{-1} \overset{\text{check}}{=} \Sigma$	$y \sim \mathcal{P}(y\vert\theta)$ $\tilde{y} \sim$ Importance sample y using $\tilde{w} = \dfrac{\mathcal{P}(y\vert\hat{\theta} + \delta)}{\mathcal{P}(y\vert\hat{\theta})}$ $\hat{\theta}(\tilde{w}) \overset{\text{check}}{=} \hat{\theta}(1) + \delta$

Future work

Student contributions and ongoing work:

- **Vladimir Palmin** is working on extending MrPlew to `lme4`
- **Sequoia Andrade** is working on generalizing to other local sensitivity checks
- **Lucas Schwengber** is working on novel flow–based techniques for local sensitivity
- **(Currently recruiting!)** Doubly–robust Bayesian MrP (the “implicit weights” path)

Vladimir Palmin

Sequoia Andrade

Lucas Schwengber

Preprint and R package coming soon! 🙏

Extra slides

Alexander, M. (2019). *Analyzing name changes after marriage using a non-representative survey*. URL: https://www.monicaalexander.com/posts/2019-08-07-mrp/.

B., Eli, Avi F., and Erin H. (2021). *Multilevel calibration weighting for survey data*. arXiv: 2102.09052 [stat.ME].

Blue Rose Research (2024). *2024 Election Retrospective Presentation*. https://data.blueroseresearch.org/2024retro-download. Accessed on 2024-10-26.

Bonica, A. et al. (Apr. 2025). *Did Non-Voters Really Flip Republican in 2024? The Evidence Says No.* https://data4democracy.substack.com/p/did-non-voters-really-flip-republican.

Chattopadhyay, A. and J. Zubizarreta (2023). “On the implied weights of linear regression for causal inference”. In: *Biometrika* 110.3, pp. 615–629.

Cohen, P. (Apr. 2019). *Marital Name Change Survey*. DOI: 10.17605/OSF.IO/UZQDN. URL: osf.io/uzqdn.

Deville, J., C. Särndal, and O. Sautory (1993). “Generalized raking procedures in survey sampling”. In: *Journal of the American statistical Association* 88.423, pp. 1013–1020.

Diaconis, P. and D. Freedman (1986). “On the consistency of Bayes estimates”. In: *The Annals of Statistics*, pp. 1–26.

Efron, B. (2015). “Frequentist accuracy of Bayesian estimates”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77.3, pp. 617–646.

Fuller, W. (2011). *Sampling statistics*. John Wiley & Sons.

G. and T. Broderick (2024). *The Bayesian Infinitesimal Jackknife for Variance*. arXiv: 2305.06466 [stat.ME]. URL: https://arxiv.org/abs/2305.06466.

G., T. Broderick, and M. I. Jordan (2018). “Covariances, robustness and variational bayes”. In: *Journal of machine learning research* 19.51.

References ii

G., W. Stephenson, et al. (2019). “A swiss army infinitesimal jackknife”. In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1139–1147.

Gelman, A. (2007a). “Rejoinder: Struggles with survey weighting and regression modelling”. In: *Statistical Science* 22.2, pp. 184–188.

— (2007b). “Struggles with survey weighting and regression modeling”. In.

Gustafson, P. (1996). “Local sensitivity of posterior expectations”. In: *The Annals of Statistics* 24.1, pp. 174–195.

Kasprzak, M., **G.**, and T. Broderick (2025). *How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences*. arXiv: 2209.14992 [math.ST]. URL: https://arxiv.org/abs/2209.14992.

Kastellec, J., J. Lax, and J. Phillips (2010). “Estimating state public opinion with multi-level regression and poststratification using R”. In: *Unpublished manuscript, Princeton University* 29.3.

Krantz, S. and H. Parks (2012). *The Implicit Function Theorem: History, Theory, and Applications*. Springer Science & Business Media.

Lax, J. and J. Phillips (2009). “Gay rights in the states: Public opinion and policy responsiveness”. In: *American Political Science Review* 103.3, pp. 367–386.

Ruggles, S. et al. (2024). *IPUMS USA: Version 15.0 [dataset]*. DOI: 10.18128/D010.V15.0. URL: https://usa.ipums.org.

Solon, G., S. Haider, and J. Wooldridge (2015). “What are we weighting for?” In: *Journal of Human resources* 50.2, pp. 301–316.

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!**

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary! Two solutions, with their own pros and cons:**

Option 1: Force $\tilde{y}$ to be binary.

Option 2: Allow $\tilde{y}$ to take generic values.

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary! Two solutions, with their own pros and cons:**

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E}\left[y|\mathbf{x}\right]$
   - E.g. Posterior mean, or
   - Shrunken posterior mean, or
   - Some values that gives the same posterior
2. Take $u_i \overset{iid}{\sim} \text{Unif}(0, 1)$
3. Assume $y_i = \mathbb{I}\left(u_i \le \hat{m}(\mathbf{x}_i)\right)$
4. Draw $u_n | y_n$
5. Set $\tilde{y}_i = \mathbb{I}\left(u_i \le \hat{m}(\mathbf{x}_i) + \delta \mathbf{x}_i\right)$

Option 2: Allow $\tilde{y}$ to take generic values.

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary! Two solutions, with their own pros and cons:**

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E}\left[y|\mathbf{x}\right]$
   - E.g. Posterior mean, or
   - Shrunken posterior mean, or
   - Some values that gives the same posterior
2. Take $u_i \overset{iid}{\sim} \mathrm{Unif}(0,1)$
3. Assume $y_i = \mathbb{I}\left(u_i \le \hat{m}(\mathbf{x}_i)\right)$
4. Draw $u_n | y_n$
5. Set $\tilde{y}_i = \mathbb{I}\left(u_i \le \hat{m}(\mathbf{x}_i) + \delta \mathbf{x}_i\right)$

Option 2: Allow $\tilde{y}$ to take generic values.

1. Set $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.
2. Then you're done.
3. There is nothing else to do.
4. This space deliberately left blank.

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\mathrm{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\mathrm{MrP}}(Y_{\mathcal{S}})$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary! Two solutions, with their own pros and cons:**

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E}\left[y|\mathbf{x}\right]$
    - E.g. Posterior mean, or
    - Shrunken posterior mean, or
    - Some values that gives the same posterior
2. Take $u_i \overset{iid}{\sim} \mathrm{Unif}(0, 1)$
3. Assume $y_i = \mathbb{I}\left(u_i \leq \hat{m}(\mathbf{x}_i)\right)$
4. Draw $u_n | y_n$
5. Set $\tilde{y}_i = \mathbb{I}\left(u_i \leq \hat{m}(\mathbf{x}_i) + \delta \mathbf{x}_i\right)$

Option 2: Allow $\tilde{y}$ to take generic values.

1. Set $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.
2. Then you're done.
3. There is nothing else to do.
4. This space deliberately left blank.

Pros and cons: (Option 1)

- Realistic
- Have to pick $\hat{m}(\mathbf{x})$
- $\tilde{Y}_{\mathcal{S}} - Y_{\mathcal{S}}$ not infinitesimally small
- **Use for checks & experiments**

Pros and cons: (Option 2)

- Not realistic
- No additional assumptions
- $\tilde{Y}_{\mathcal{S}} - Y_{\mathcal{S}}$ may be infinitesimally small
- **Use for theory**

Real Data: Lax Philips

Analysis of national support for gay marriage.[13]

- **Target population:** US Census Public Use Microdata Sample 2000
- **Survey population:** Combined national-level polls from 2004
- **Respose:** "Do you favor allowing gay and lesbian couples to marry legally?"
- For regressors, use race, gender, age, education, state, region, and continuous statewide religion and political characteristics, including some analyst–selected interactions.

Survey observations:	$N_S = 6,341$
Target observations (rows):	$N_T = 9,694,541$

Uncorrected survey mean:	$\frac{1}{N_S}\sum_{i=1}^{N_S} y_i = 0.333$
Raking:	$\hat{\mu}_{\text{WGT}} = 0.33$
MrP:	$\hat{\mu}_{\text{MrP}} = 0.337 \quad (\text{Post. sd} = 0.039)$

[13] Based on Kastellec, Lax, and Phillips (2010), see also Lax and Phillips (2009).

Covariate balance for primary effects

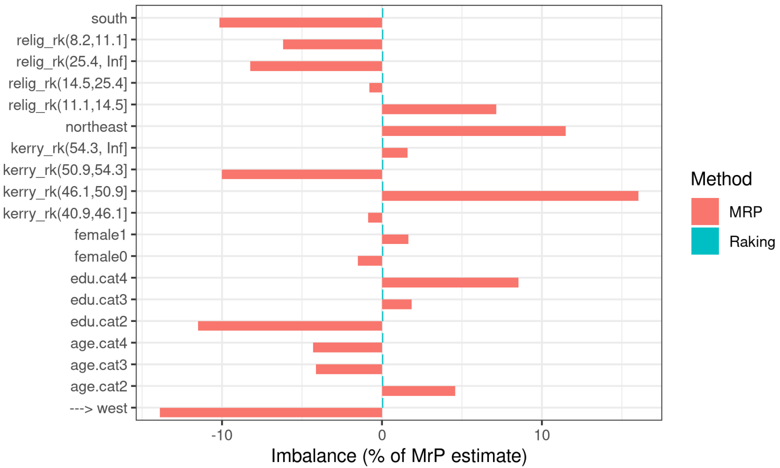


Figure 8: Imbalance plot for primary effects in the Gay Marriage dataset

Covariate balance for interaction effects

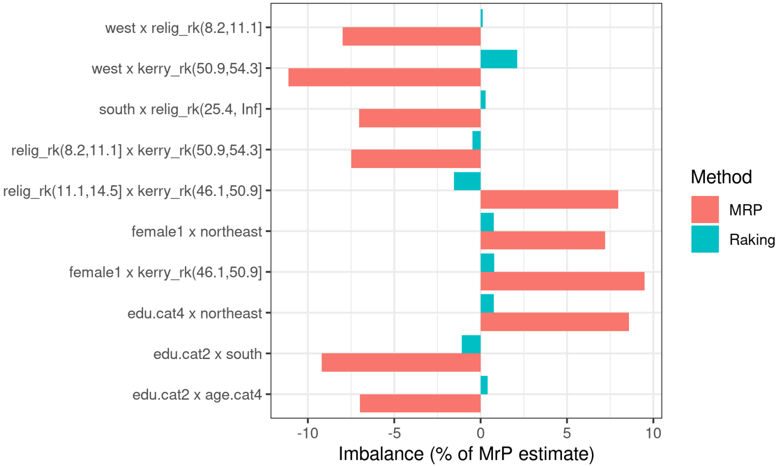


Figure 9: Imbalance plot for select interaction effects in the Gay Marriage dataset

Predictions

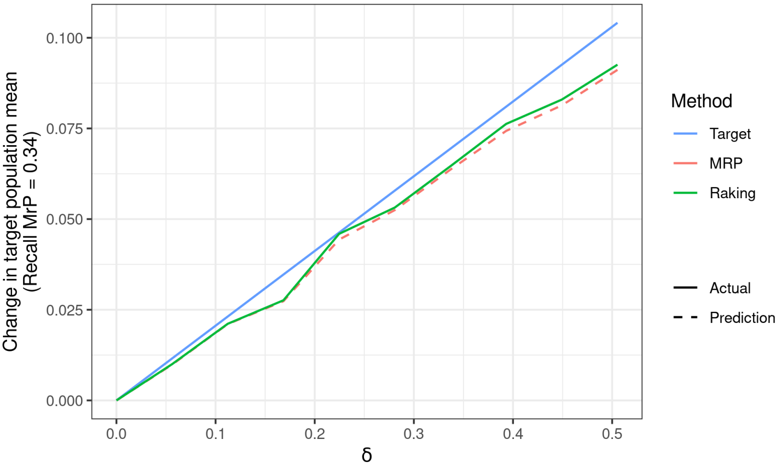


Figure 10: Predictions on binary data for the Gay Marriage dataset

Predictions and actual MCMC results

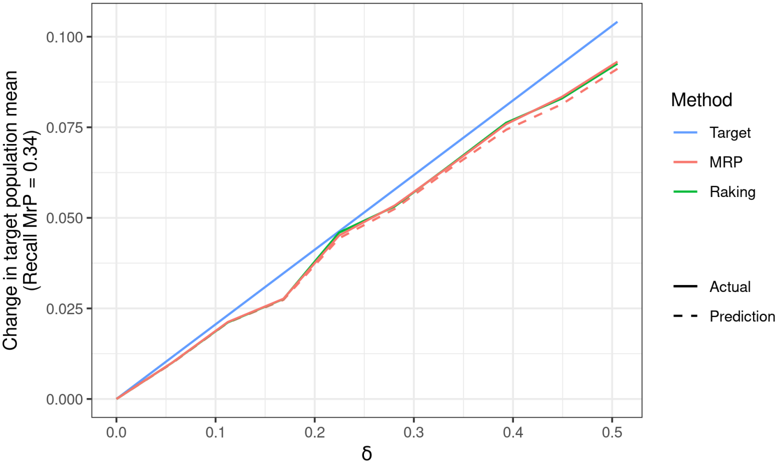


Figure 11: Predictions and refit on binary data for the Gay Marriage dataset

Running ten MCMC refits: 11 hours Computing approximate weights: 23 seconds